DOCUMENT RESUME

ED 313 640 CG 022 163

AUTHOR Krug, Samuel E.
TITLE Solid State Psychology: The Impact of Computerized Assessment on the Science and Practice of Psychology.
PUB DATE 13 Aug 89
NOTE 15p.; Paper presented at the Annual Meeting of the American Psychological Association (97th, New Orleans, LA, August 11-15, 1989).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computers; *Measures (Individuals); *Psychological Evaluation; Psychological Services; *Psychological Testing
IDENTIFIERS *Computer Assisted Counseling

ABSTRACT

Computer-assisted testing is not without its problems and pitfalls, but it holds a great deal of promise as well. Computer administration of tests provides more control over the testing process than was ever possible with paper-and-pencil testing. At the same time it offers the possibility of being able to monitor and record aspects of the testing process, such as response latency and response shifting, that may prove to be important predictive factors in their own right. Computer scoring of tests has made it possible to obtain accurate scores. It has been estimated that errors involving a difference of one or more points in the final score are made in 10 percent of cases involving hand scoring of objective tests. These and similar errors of measurement may have more impact on the reliability of scores obtained in practice than some of the better analyzed sources described in measurement theory. In the final analysis, computer interpretations of test scores may offer the greatest potential for advancing psychological measurement. As the volume of research data relevant to a particular test increases, the task of using it effectively in interpretation becomes increasingly frustrating for the unassisted test user. Perhaps even more importantly, computerized reports produce consistent, predictable outputs that can be analyzed and improved if the appropriate models and techniques are developed for doing so and they are treated scientifically and not as scientific curiosities. (ABL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED313640

SOLID STATE PSYCHOLOGY: THE IMPACT OF COMPUTERIZED ASSESSMENT ON THE SCIENCE AND PRACTICE OF PSYCHOLOGY[1]

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Samuel E. Krug
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Samuel E. Krug
MetriTech, Inc.
Champaign, Illinois

There are very few activities which have not been profoundly altered by rapid advances in computer technology. Psychology is certainly no exception to that rule. A recent national survey of 1312 psychologists, social workers, and marriage/family counselors reported that almost 60% now own a computer. Another 15% say they plan to buy one in the near future (Psychotherapy Finances, 1988).

Obviously, many do so for reasons unrelated directly to the practice of psychology (e.g., billing software, word processing, etc.). However, the administration, scoring, and analysis of test results, which occupies a significant portion of many psychologists' time, has become heavily computerized.

During the next hour, I would like to do the following three things.

First, I would like to review briefly the introduction of the computer into the assessment process and the variety and quantity of CATPs currently available.

Next, I will suggest that these products that psychologists purchase and use in rapidly increasing numbers vary considerably in quality, particularly with respect to computerized narrative reports and that no good standards yet exist for evaluating them.

Finally, I will suggest that improvements in such products can only come about when we begin to develop quantitative theory and formal statistical criteria for evaluating them.

THE ROLE OF THE COMPUTER IN PSYCHOLOGICAL ASSESSMENT

The primary purpose of psychological testing is to transform individual characteristics into numbers that can be used to make more reliable and valid decisions about individuals. As our knowledge of psychometrics increased, the assessment process itself underwent considerable transformation. What was once simple, has become increasingly complex. For that reason and others, reliance upon the computer in assessment became critical.

The introduction of computers into assessment was innocent enough. In the beginning was the scoring machine. Even before digital computers as we know them appeared, their ancestors—key punches, card sorters, and listing machines—were pressed into service. Moreland (1987) reports that in the late 1920s the 22 scores of the Strong Vocational Interest Blank (SVIB) could be obtained by passing 420 Hollerith cards through a sorter several times. Considering the alarming frequency with which sorters "ate" cards, this was not an activity for the weak-willed. However, it serves as a dramatic illustration of the lengths to which people will go in order to avoid hand scoring.

---

[1] Invited address to the 97th Annual Convention of the American Psychological Association. New Orleans, August 13, 1989.

CG 022163

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

As computers became more powerful, more available, and more economical, the contribution they made to assessment increased dramatically. For example, modern test construction would be virtually impossible without computers. Those who were committed to broadening the scope of assessment looked for ways in which the computer could be of further assistance. And so, it was not altogether unexpected when machines that were used to score and develop these instruments began to administer and interpret results as well.

THE VARIETY OF PRODUCTS AVAILABLE

Since computer-assisted test products (CATPs) first appeared little more than a quarter century ago, the domain has expanded rapidly to encompass a wide array of applications. Hundreds of products have been designed for clinical diagnosis, educational evaluation, marital counseling, and career development, for example (Krug, 1986; 1987a).

For several years, I've been charting this area fairly carefully and regularly for *Psychware* (Krug, 1984, 1987b, 1988). The first edition of this guide to computer-assisted test products, published in 1984, contained descriptions and samples for 191 entries. The second edition, published in early 1987 included 339 entries. The third and current edition, published one year ago, included a total of 451 entries, more than twice as many as were included in the first edition. Figure 1 provides a graphic breakdown of products by category[2]. Figure 2 provides a breakdown of products by application[3].

Insert Figure 1

With only a few notable exceptions the categorization of products has remained relatively stable throughout the last half decade. The number of products for use in neuropsychological assessment (NP) increased substantially between 1984 and 1987. The large increase in the number of Utility (UT) products in the third edition of *Psychware* is partly artifactual. In earlier editions, entries tended to be identified with a single test or assessment procedure. For the third edition in 1988, we broadened the definition of what could be included in *Psychware*.

Insert Figure 2

As Figure 2 shows, the classifications have also remained stable from 1984-1989. By far, the largest number of products are designed for Clinical Assessment or Diagnosis (C). Products for Individual Counseling (PC) and Vocational Guidance (GC) are next, followed closely by products for Personnel Selection and Educational Evaluation (SE). Products for use in Behavioral Medicine (BM) have essentially tripled from 1984 to 1988, although they still represent a relatively small proportion of product applications.

---

[2] For Figure 1 the categories are defined as follows: CV-Career/Vocational; AC-Ability/Cognitive; IA-Interests/Attitudes; M-Motivation; NP-Neuropsychological; P-Personality; SI-Structured Interview; UT-Utility. See Krug (1988, p. xv) for additional description of these categories.

[3] For Figure 2 the applications are defined as follows: BM-Behavioral Medicine; C-Clinical Assessment/Diagnosis; EE-Educational Evaluation/Planning; PC-Individual Counseling; LD-Learning Disability Screening; MF-Marriage/Family Counseling; SE-Personnel Selection/Evaluation; TD-Training/Development; GC-Vocational Guidance/Counseling. See Krug (1988, p. xvi) for additional description of these categories.

Considering the rapid growth reflected in these two tables, the following conclusion seems inescapable: after little more than a quarter century, computer testing has become an overnight sensation.

CONCERNS ABOUT THE QUALITY OF CATPs: HOW GOOD ARE THEY?

At first glance, it would appear we face what the French describe as an "embarrassment of riches." And, at first glance, that would seem to be a very reasonable conclusion. Many of the products currently available are well crafted tools that make important contributions to the decision-making process. Consider results from a recent survey of 329 users of computer-assisted test products which yielded a total of 576 individual product ratings. As you might expect of user surveys, the ratings tended on the whole to be skewed. For example, on a five-point Likert-type scale, with lower ratings reflecting greater dissatisfaction with the product, the average rating was 3.89.

Since the overall rating was itself an average of several items that were independently rated, it was possible in this study to identify elements that contributed most to overall satisfaction/dissatisfaction. As Table 1 shows, the usefulness of the information the product provides received the highest overall rating. At the bottom end of the table, it would appear that computer-assisted test products live very much in their own world. That is, they do not easily integrate with other computer programs.

Insert Table 1

Table 1 shows not only the average ratings across product by item but also the variance across product ratings explained by each item. As you will note, more than a fourth of the (population) variance in product ratings is attributable to the item "How well documented is the development of this product." This may explain why so many have been concerned less about the riches these products provide and more about the scientific embarrassment they may represent.

Concerns about the potential that exists for misusing computer-assisted test products have been heard from many different quarters (for example, Eyde and Kowal, 1985; Matarazzo, 1983; Mitchell, 1986). Two concerns that are often mentioned are: 1) the depersonalization of the assessment process and 2) the technical quality of the products being offered.

Depersonalization

Depersonalization is an especially important concern when human services are involved. Some feel that the computer increases the distance between the service provider and client, leading to decreased communication and a more mechanized, but less effective delivery system. Based on published research, however, this would appear to be more a concern of the therapist than the client. Skinner and Allen (1983) and Harrell and Lombardo (1984) have suggested that clients actually prefer to face the computer than a live interviewer or a test booklet. And Wagman's research (1980, 1982) suggests: 1) that computerized counseling results in about the same gains as are found for more traditional approaches and 2) that clients often prefer the computer to a live therapist.

In reality, many other human service systems have had to permit some degree of depersonalization in order to take advantage of more effective diagnostic and treatment techniques. For example, before radiology, a fracture could only be crudely diagnosed by sight

and touch. Patients have had to trade some of the patient-physician relationship they formerly enjoyed to take advantage of these new techniques. But the overall effect has been to improve the effectiveness of medical practice and the quality of life itself.

A reasonable conclusion appears to be that the introduction of technology is not in itself depersonalizing. In fact, quite the opposite may be true. For example, the use of the computer may free the practitioner from routine tasks, such as administering, scoring, and analyzing test results leaving more time available for interaction with clients.

Technical Quality of the Products Being Offered

For several years now, the technical quality of computer-generated test reports has been a subject of particular interest. These products range in complexity from very straightforward score reports to expansive computer-generated narratives. Zachary (1984), for example, has distinguished five major classes or types of products: scoring reports, extended scoring reports, descriptive reports, screening reports, and consultative reports. Most consist of a combination of numeric, graphic, and narrative elements. However, it is the narrative component that has aroused the greatest concern among professionals.

Numerous writers have recognized the need to validate the interpretive component of computerized report systems (Eyde, Kowall & Fishburne, 1990; Eyde & Kowall, 1989; Moreland, 1987). Many studies involve so called "customer satisfaction" designs that ask users to rate the "accuracy" of the narrative descriptions. The problem is that accuracy is often loosely defined. Sometimes it is taken to mean "precise," a characteristic usually associated with reliability in the case of test scores. Other times it is taken to mean "correct," a characteristic associated with test validity. When it is loosely defined, accuracy can be easily confounded. For example, some systems may consist largely of high base rate statements that are true for 95% of the population. In the same way that a test score may be reliable without being valid, such systems may be "accurate" without being useful.

In addition to this problem, most accuracy studies have operated at a very global level of analysis. Moreland (1987, p. 42), for example, summarized results of 15 studies that dealt with the accuracy of five MMPI systems. In each study, the primary outcome variable was an overall accuracy index that ranged from 32% to 85% within one system and from 32% to 91% across all five systems. The use of such an index may be very appropriate for comparing various systems or market surveys. However, such a level of analysis is not helpful in identifying "inaccurate" elements of a single system nor in improving the technical quality of products in general. How comfortable would we be in selecting tests if manuals reported only a single "accuracy" index as the sum total of evidence offered in support of the test?

In many ways the current state of analysis and development in the field of computer-generated test reports resembles the state of measurement itself at the turn of this century. Scientists like Galton in England and J. McKeen Cattell in the United States, who regarded the measurement of individual differences as a better way of developing laws of human behavior, failed in some of their earliest attempts to establish systematic relationships between test scores and socially significant outcomes. For example, correlations between Cattell's "mental tests" and college grades were disappointingly and uniformly low, the highest correlation being .19 (Gulliksen, 1950). These individual differences pioneers relied too heavily on an assessment methodology inherited from the physical science laboratories where the only source of error was thought to lie in the observer, not the observation.

Spearman and others soon recognized that a careful study of test characteristics was a necessary prerequisite for any real advancement in psychology. In fact, it was Spearman's

(see, for example, Spearman, 1904; 1907) introduction of quantitative theory and mathematical models to describe the structure and behavior of test scores that opened the door to modern psychometrics.

Cronbach (1984) defined a test as a systematic procedure for recording and *describing* (emphasis added) behavior through the use of numerical scales or fixed categories (p. 26). In a very important sense, the interpretation of the test profile and the resulting behavioral description is as much a part of the test as the items, keys, and norms. The real promise of computer technology does not lie in the fact that the computer can write reports faster or more economically than a single clinician. It lies instead in the fact that computer-based test reports offer the potential of being able to produce better, more valid, and more useful interpretations than a single clinician. However, this promise will not be realized until we begin to develop a unified theory that can advance the science of computer-generated reports in the same way, for example, that classical test theory and item response theory have advanced the science of observation.

Concepts, Models, and Methodologies for Evaluating Computerized Narrative Reports

A comprehensive discussion of concepts, models, and statistical criteria for describing and evaluating the structure and behavior of computer-generated narratives is beyond the scope of the present paper. However, a brief presentation along these lines may stimulate thinking about the formal nature of such systems among those interested in their use and evaluation.

Just as a test is composed of discrete *items*, a computerized narrative may be thought to be composed of a series of discrete *inferences*. In some cases, an inference may be represented by a phrase or a single sentence. In other cases, an inference will encompass an entire paragraph or set of paragraphs. Generally, the definition of inferential elements within a report will be made by the system author and will correspond operationally to output associated with a single decision rule or set of rules. That is, in some cases a phrase, sentence, or paragraph will be associated with a score on one test scale (single rule). In other cases, the same narrative may be produced by several, alternative profile configurations.

Following developments within classical test theory, we may further suppose that each inference includes some element of error. That is, any narrative contains some information about a person that is true and some information that is not true. Although these two components might at first glance appear to be directly calculable, like true scores and error scores they are actually unknown quantities that can be only indirectly estimated. For example, since an examinee theoretically has access to a broader sampling of observations on which to base a judgment, he or she may decide that the statement "The client is very likely to become upset or agitated in situations that require him to work closely with others" is correct. Someone else who has seen that client only in a one-on-one situation may decide that the statement is wrong.

In classical test theory, the error associated with observations is called error of measurement. One goal of test theory is to identify the sources and magnitudes of such errors so as to minimize their impact. This is done by analyzing test scores and the items that contribute to test scores under various experimentally defined conditions. Some items contribute undesirably large proportions of error variance to the total score. The item analysis process allows us to identify those items and eliminate them before a test is released for operational use.

With respect to computer-generated narrative reports, we may define a comparable term, error of interpretation, as the error associated with narrative inferences. By systemat-

ically studying the performance of computer-generated narrative reports and their component inferences under carefully controlled conditions, it should therefore be possible: 1) to make summary statements about relevant characteristics of a narrative report as a whole and 2) to identify formal statistical criteria by which to evaluate the contribution and appropriateness of individual components. These two objectives may be seen to parallel reliability analysis and item analysis, respectively, in the case of test scores. The purpose of such studies, of course, is to reduce error sources and produce potentially more valid and valuable products.

Reliability of Computer-Generated Narratives

With respect to tests, reliability studies are designed to identify and measure various sources of error that affect the precision of scores. A test is said to be reliable to the extent that individuals obtain similar scores across changes in conditions, such as administrators, scorers, time, or sets of items thought to be parallel. In a broader sense, reliability may be described as the extent to which scores replicate or generalize in anticipated ways from one observation to another (Cronbach, Rajaratnam, & Gleser, 1963). Generalizability theory with its concept of facets of observations and associated statistical designs may represent a productive way to begin looking at computer-generated reports.

There are numerous conditions which may be of potential interest to report users. First, consider the generalizability of reports across time, a concept analogous but not identical to that of test-retest reliability. Although the stability of a narrative is correlated with the stability of the underlying test, it is not statistically dependent on it. For example, one system author may decide to report a single statement over a broad range of test scores. Under such circumstances fluctuations in narrative content are likely to be less than fluctuations in test scores. On the other hand, an author may try to build a great deal of variety into the statement library so that even relatively minor score differences trigger different output.

The concept of generalizing across inferences is analogous to generalizing across test items, a characteristic that reflects upon the internal consistency of a test. This is often a desirable quality of tests, at least when the content domain is thought to be unidimensional. On the other hand, with respect to computer-generated narratives, the expectation may be very different. Is it the case, for example, that inferences are essentially redundant, which would be reflected in a high index across inferences, or is the system sensitive to differences within the individual, which would be reflected in a low index across inferences?

An Illustration

This type of study requires only a two factor design—persons x inferences—and is exactly analogous to evaluating the internal consistency of a set of test items. The basic data matrix consists of rows of persons by inference "scores." An inference score has only two values: it is "1" if the inference appears in the person's report and "0" otherwise.

The narrative report for the microcomputer version of the Adult Personality Inventory (Krug, 1985) was designed to focus primarily on elements of the test profile that were distinctive, not common. That is, the intent was to produce a short narrative composed of relatively low base-rate statements. Table 2 presents results of an empirical study conducted to illustrate the type of design described here.

Insert Table 2

As Table 2 shows, there are 27 possible narrative inferences in this particular computer generated report. Using variance estimates from the ANOVA summary table, we find that the internal consistency of the report is very low (.05). That is, there is very little correlation between one inference and another. Or, to put in another way, individual inferences do not appear to be drawn from the same universe. Although this would be a problem if we were talking about items in a mathematics test, the finding is consistent with the design objectives of the narrative, that is to report only on distinctive features of the person, not features that are likely to typify many people.

With other products, alternative outcomes might be preferable. For example, in a report consisting of inferences nested within homogeneous topic areas, one would expect to find a similar degree of differentiation across topic areas, but a higher degree of internal consistency within topic area. The point is that it is not necessarily desirable to have high or low values. Rather, it is desirable that the values match the design specifications of the system author and the purposes of the report user.

Earlier I mentioned that a quantitative approach to report analysis would be helpful also in analyzing individual inferences in a manner similar to item analysis. One way in which this might be done is illustrated by the data reported in the second half of Table 2. This shows the correlation of each inference with the inference total score and the alpha coefficient if this inference were to be removed from the report. Keeping in mind that the design objectives of this particular report, low consistency is desirable. Consequently, these data indicate that Statement 27 is most helpful in meeting the design objectives and Statement 19 is least helpful. Just as item analysis information helps a test author refine the test, quantitative information of this sort can be helpful to the report author in refining the decision rules that produce each statement.

Obviously, there are many other designs and statistics that need to be considered in the evolution of what might be called "classical report theory." My comments today are intended only to stimulate thinking along such lines and to suggest that a more quantitative approach to the study of report narratives may return significant dividends for both system authors and system users.

SUMMARY

Computer-assisted testing is not without its problems and pitfalls. But it holds a great deal of promise as well.

Computer administration of tests provides more control over the testing process than was ever possible with paper-and-pencil testing. At the same time it offers the possibility of being able to monitor and record aspects of the testing process, such as response latency and response shifting, that may prove to be important predictive factors in their own right.

Computer scoring of tests has made it possible to obtain accurate scores. Gorsuch (personal communication, August 17, 1988) has estimated that errors involving a difference of one or more points in the final score are made in 10% of cases involving hand scoring of objective tests. These and similar "errors of measurement" may have more impact on the reliability of scores obtained in practice than some of the better analyzed sources described in measurement theory.

In the final analysis, computer interpretations of test scores may offer the greatest potential for advancing psychological measurement. As the volume of research data relevant to a particular test increases, the task of using it effectively in interpretation becomes increasingly frustrating for the unassisted test user. Perhaps even more importantly, computerized reports produce consistent, predictable outputs that can be analyzed and improved if we develop the appropriate models and techniques for doing so and begin treating them scientifically, not as scientific curiosities.

REFERENCES

Cronbach, L. J. (1984). *Essentials of psychological testing (4th ed.)*. New York: Harper & Row.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1968). Theory of generalizability: A liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology, 16*, 137-163.

Eyde, L. D., & Kowal, D. M. (1985, August). Psychological decision support software for the public: Pros, cons, and guidelines. Paper presented at the 1985 Convention of the American Psychological Association, Los Angeles.

Eyde, L. D., Kowall, D. M., & Fishburne, F. J. (1990). The validity of computer-based test interpretations of the MMPI. In S. Wise and T. B. Gutkin (Eds.), *The computer as adjunct to the decision making process*. Lincoln: University of Nebraska.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Harrell, T. H., & Lombardo, T. A. (1984). Validation of an automated 16PF administration procedure. *Journal of Personality Assessment, 48*, 638-642.

Krug, S. E. (1984). *Psychware*. Kansas City: Test Corporation of America.

Krug, S. E. (1985). *TEST PLUS: A microcomputer based system for the Adult Personality Inventory*. Champaign, IL: MetriTech.

Krug, S. E. (1986). Solid state psychology: The role of the computer in human assessment. In R. B. Cattell & R. C. Johnson (Eds.), *Functional psychological testing: Principles and instruments*. New York: Brunner/Mazel.

Krug, S. E. (1987a). Microtrends: An orientation to computerized assessment. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide*. New York: Basic Books.

Krug, S. E. (1987b). *Psychware sourcebook: 1987-1988, Second edition*. Kansas City, MO: Test Corporation of America.

Krug, S. E. (1988). *Psychware sourcebook: 1988-1989, Third edition*. Kansas City, MO: Test Corporation of America.

Matarazzo, J. D. (1983). Computerized psychological testing. *Science, 221*, 323.

Mitchell, J. V. (1986). Measurement in the larger context: Critical current issues. *Professional Psychology: Research and Practice, 17*, 544-550.

Moreland, K. L. (1987). Computerized psychological assessment: What's available. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide*. New York: Basic Books.

Staff (1988). Testing in private practice. *Psychotherapy Finances, 15(3)*, 7.

Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology, 51*, 267-275.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161-169.

Wagman, M. (1980). PLATO DCS: An interactive computer system for personal counseling. *Journal of Counseling Psychology, 27*, 16-30.

Wagman, M. (1982). Solving dilemmas by computer or counselor. *Psychological Reports, 50*, 127-135.

Zachary, R. (1984, August). Computer-based test interpretations: Comments and discussion. In J. D. Matarazzo (Chair), *Computer-based test interpretation: Prospects and problems*. Symposium conducted at the annual convention of the American Psychological Association, Toronto.

Figure 1

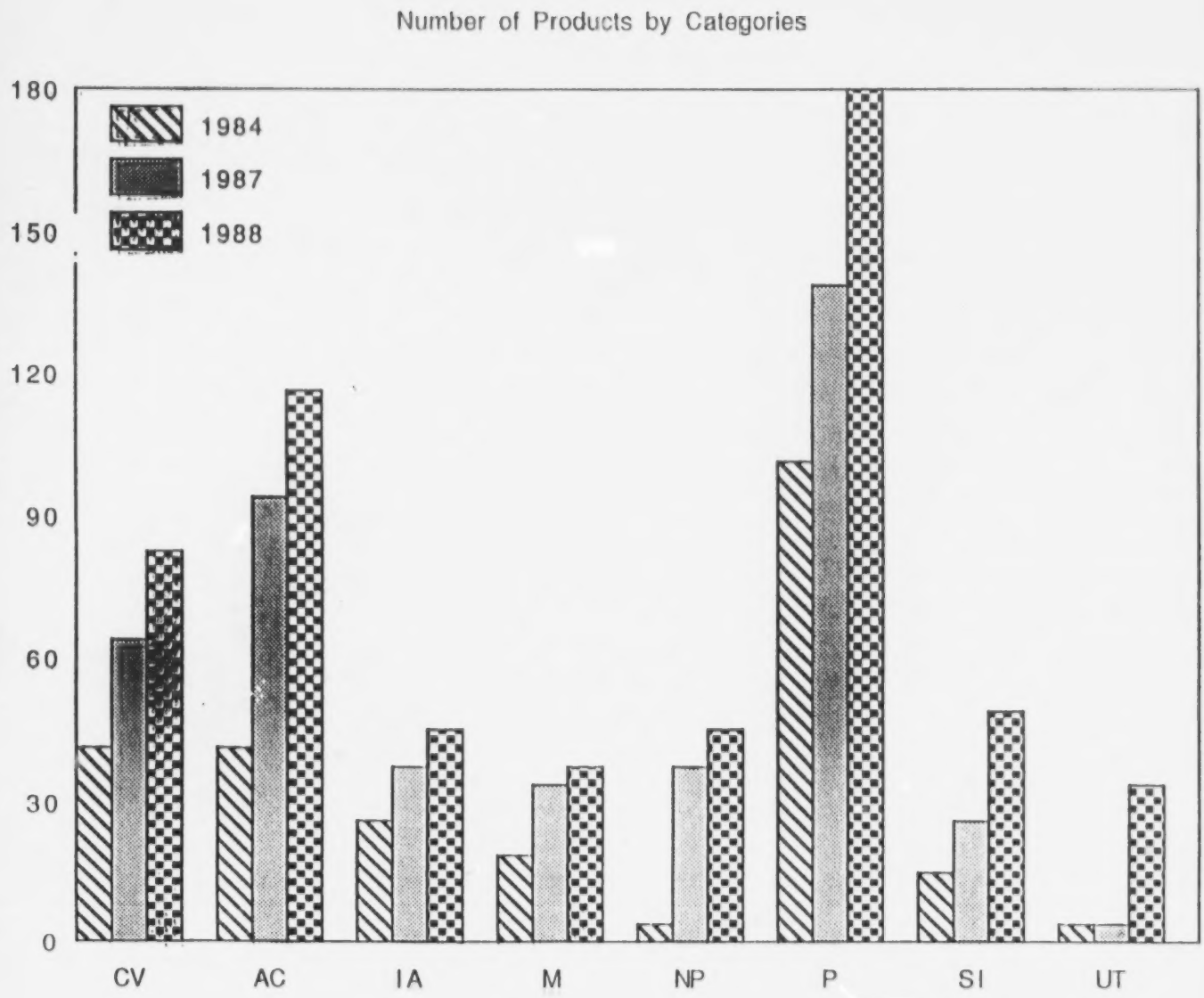
Number of Products by Categories
180
150
120
90
60
30
0
1984
1987
1988
CV
AC
IA
M
NP
P
SI
UT

Figure 2

Number of Products by Application

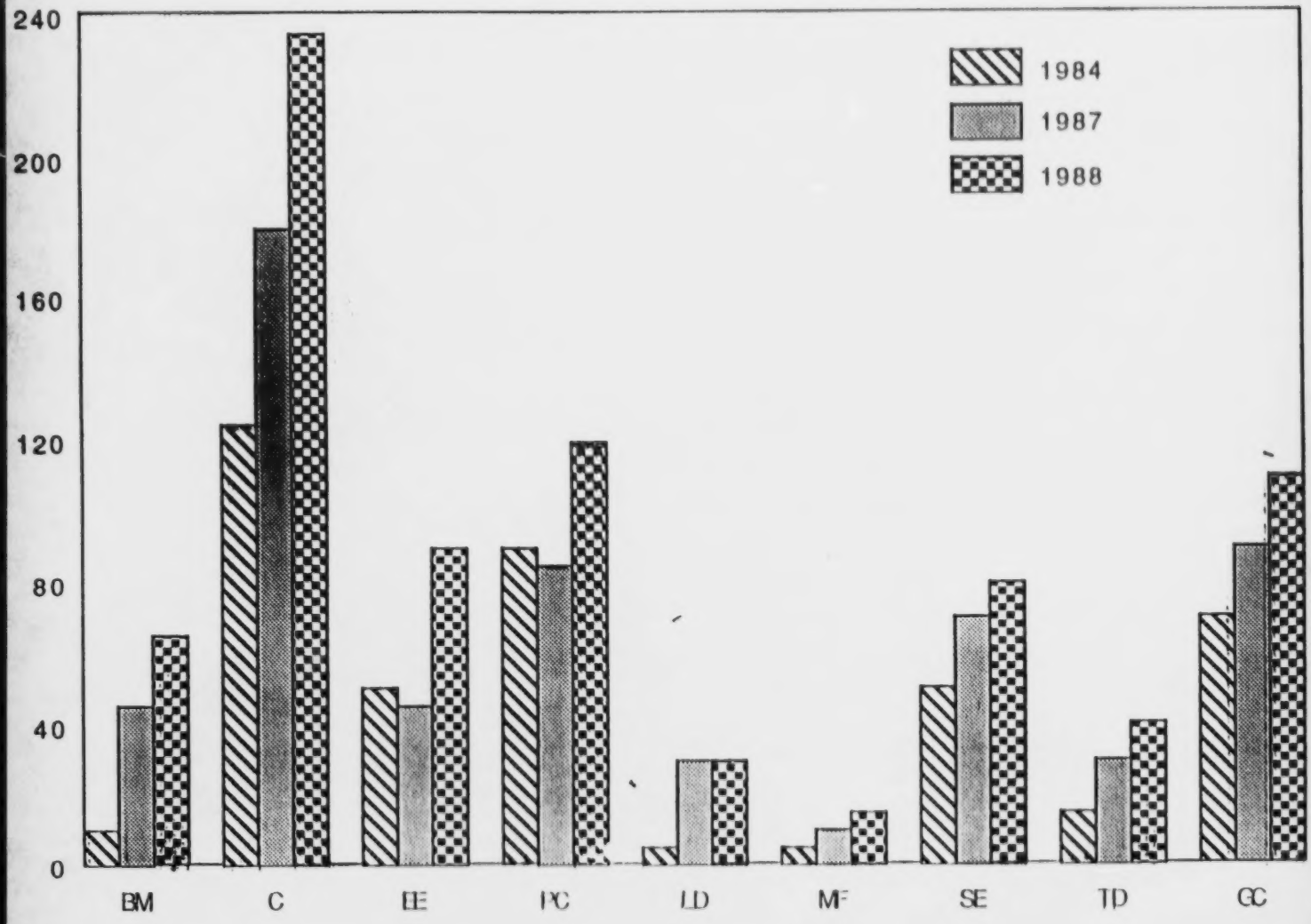

Table 1

Results of Computer-Based Product Rating Study

Average Rating	Item	Variance Explained Across Products
4.29	How useful is the information this product provides?	.11
4.24	Overall, how easy is it to use this product?	NS
4.20	How frequently do you encounter problems in using this product?	NS
4.10	How well does this product use computer technology?	.13
4.08	Overall, how valuable or cost-effective is this product?	.12
4.03	How good is the quality of overall support provided by the supplier?	.12
4.00	How often do the results from this product conflict with your professional judgment?	.10
3.95	How well documented is the development of this product?	.26
3.89	How much does the information from this product enhance your professional decision making?	.06
3.83	How helpful are the user's manuals or instructions?	.13
3.22	How much training or study is required to use this product effectively?	.12
2.55	How easily does this product integrate with other computer programs you use?	.06

Based on a total of 576 ratings of 121 computer-assisted test products provided by 329 raters.

Table 2

Results of Generalizability Study Across Inference Elements of the Adult Personality Inventory Narrative Report

Analysis of Variance

Source	Sum of Squares	df	Mean Square	F
Persons	98.951	557	.178	1.053
Inferences	226.252	26	8.701	51.485
P X I	2445.155	14,482	.169	
Total	2770.358			

Inference	Corrected Item-Total Correlation	Alpha if Deleted
1	.1368	.0011
2	-.1003	.0920
3	.0557	.0398
4	.0057	.0497
5	-.0554	.0697
6	.0539	.0304
7	.1352	-.0094
8	.0358	.0397
9	-.1246	.0997
10	.0908	.0170
11	-.0546	.0771
12	-.0196	.0540
13	-.1057	.0859
14	.0035	.0497
15	-.0191	.0593
16	.0258	.0406
17	.1159	-.0008
18	.2260	-.0299
19	.2679	-.0493
20	.0702	.0293
21	-.2072	.1253
22	.1668	-.0231
23	.1344	-.0119
24	-.1506	.1093
25	-.0774	.0875
26	-.0265	.0611
27	-.2934	.1353

Based on data from 279 men and 279 women.